Method Validation

Lebah Lugalia
QA/QC Lab Coordinator
UNC Project
Lilongwe, Malawi

Method Validation

- What is method validation?
- Why is it important to validate a method?
- What regulations and guidelines apply?
- What assays must be validated?
- What experiments are involved?
- What criteria is available for method acceptance?

History of Validation

- The concept originated in analytical chemistry to verify that a method provided an accurate and representative value for the sample employed under the conditions used.

Method Validation – What is it?

- VALIDATION = ERROR ASSESSMENT
- Estimation of how much error might be present in a test result produced by a method in your laboratory.

Method Validation – Why is it important?

- You want to validate the manufacturer’s claims for their method performance characteristics.
- You want to ensure that the amount of error of the method won’t affect the interpretation of the test result and compromise patient care.

Method Validation- Why is it necessary to validate a new method

- Method performance is affected by many factors:
  - Changes in manufacturing from the production of prototypes to final field instruments
  - Effect of shipment and storage
  - Local climate conditions in your lab e.g. temp, humidity
  - Quality of water
  - Stability of electrical power
  - Skills of the operators.

Method Validation: Why is it necessary to validate a new method

- Method validations provides assurance that a new method, with whatever changes that may have occurred, still performs acceptably under the conditions of use in your laboratory.

What are the regulations?

- On Jan 24, 2003 the Centers for Medicare Services (CMS) published the Final CLIA Rule. It states that…
- “Beginning April 24, 2003 laboratories introducing non-waived methods must validate the methods performance.”
- If using an FDA approved unmodified method, you must demonstrate that your lab can obtain performance specifications comparable to those established by the manufacturer for:
- Accuracy, precision, reportable range of test results and…
- Verify that the manufacturer’s reference intervals (normal reference ranges) are appropriate for the lab’s patient population

What are the regulations?

- CLIA regulations are based on the complexity of the test method.
- Test methods are classified into 3 categories:
- 1. Waived tests
- 2. Non-waived unmodified tests – moderate and high complexity tests
- 3. Non-waived, modified (or in-house developed) tests.
- Information about classification of specific tests is available at : http://www.cms.gov./clia/

What are the regulations

- Waived Tests: tests that are simple to perform.
  - e.g. urine dipstick testing, urine pregnancy testing, HIV rapid tests, fecal occult blood tests, HemaCue Hgb test etc.
  - Method validation is not required
  - Follow manufacturer's directions for testing.

What are the regulations?

- Non-waived, unmodified tests. (Moderate and High Complexity) e.g. chemistry, hematology etc
- Validation is accomplished by performing 4 experiments:

  1. Linearity experiment (reportable range)
  2. Replication experiment (estimate imprecision)
  3. Comparison of methods (estimate inaccuracy)
  4. Establish reference intervals.

What are the regulations?

- Non-waived tests modified or developed in-house, one must determine the following performance characteristics:

  - Accuracy
  - Precision
  - Reportable range
  - Reference Intervals
  - Analytical sensitivity (detection limit)
  - Analytical specificity (interfering substances)

Accuracy

- A measurement of the exactness of an analytical method, or the closeness of agreement between the measured value and the true value.
- Inaccuracy = systematic error

Inaccuracy = Systematic Error

- Usually quantified by comparing a method to a “gold standard”
- Compare value between the “test method” and the “gold standard” to estimate the SE
- Systematic error may stay the same over a range of values or may change as concentration changes.

Precision

- Defined is the degree of agreement among individual test results obtained when the procedure is applied repeatedly to multiple samplings of a homogeneous sample.

Imprecision = Random Error

- Defined as an error that can either be positive or negative, whose direction and exact magnitude cannot be predicted.
  - Usually quantified by the standard deviation (SD).
  - SD usually increases as concentration increases
    - Therefore it is useful to calculate the coefficient of variation (CV%), which expresses the error as a percentage of the mean concentration.

Total Error (TE)

- Defined as the net or combined effect of random and systematic errors:
  - TE = RE + SE

Total Error

- Regulatory agencies, define acceptable error in terms of "total allowable error" (TE_a)
  - e.g., CLIA:

ALT:	target value +/- 20%
Potassium	target value +/- 0.5mM/L
Albumin	target value +/- 10%
Hemoglobin	target value +/- 15%
Magnesium	target value +/- 25%
Leukocyte count	target value +/- 15%

  - Listing of total allowable errors from CLIA:
    - www.westgard.com/clia.htm

Method Validation- Factors to Consider

- Factors to consider:
- Define a quality requirement for the test in the form of the amount of error that is allowable.
- Make a plan and write an outline for each validation experiment.
- Schedule ample time to perform the experiments.
- Familiarize the techs with the validation experiments.
- Make sure the instrument/method is functioning properly. i.e. is passing qc and calibration.
- Enough reagents and supplies in stock.

Method Validation

- **Replication Experiment:**
- A replication experiment is performed to estimate the imprecision or random error of the analytical method.

Replication Experiment

- Imprecision or random error is caused by:
  - Pipetting of samples
  - Reaction conditions (timing, mixing, temperature, heating,)
  - Measurement itself
  - Operator technique
  - The instability of the instrument

Replication Experiment

- Factors to consider:
  - Time period
    - Within-run/ within day measurements
    - Between-day measurements (over $\geq$ 20 days).
  - Sample selection
    - Standard solutions
    - Control Solutions
    - Pools of fresh patient samples
  - Number of samples to be analyzed

Replication Experiment – Minimum Studies

- Select at least 2 different control/standard materials or patient specimens that represent low and high medical decision concentrations for the test of interest.
  - Analyze each material 20 times within a run or within a day
    - Short-term imprecision/random error
  - Analyze each material once per day for 20 days
    - Long- term imprecision/random error

Replication Experiment

- For each of the 20 test results obtained from a single source material:
  - Calculate the Mean, SD, and CV%
  - An internet calculator is available at http://www.westgard.com/mvtools.html.

Replication Experiment - Example

- Analyte: **ALB** Method: **Synchron CX-5** Materials: **Synchron Control Levels 1,2,&3**
- Serial No.: **7244**

#	Date	Time	Tech Init.	Control L-1	Control L-2	Control L-3	Comments
1	Oct. 25, 04	10:57	HK	2.2	3.5	5.0	
2	Oct. 25, 04	16:07	HK	2.2	3.7	5.0	
3	Oct. 26, 04	9:00	HK	2.2	3.6	5.0	
4	Oct. 26, 04	16:10	HK	2.2	3.6	5.1	
5	Oct. 27, 04	11:49	HK	2.2	3.6	5.0	
6	Oct. 27, 04	14:33	HK	2.2	3.5	5.0	
7	Oct. 28, 04	9:16	HK	2.2	3.6	5.0	
8	Oct. 28, 04	13:07	HK	2.2	3.6	5.0	
9	Oct. 29, 04	9:27	HK	2.2	3.6	5.0	
10	Oct. 29, 04	15:11	HK	2.2	3.5	4.9	
11	Nov.1, 04	12:34	HK	2.2	3.6	4.9	
12	Nov.1, 04	14:29	HK	2.1	3.6	5.0	
13	Nov.2, 04	8:43	HK	2.1	3.5	5.0	
14	Nov.2, 04	14:49	HK	2.1	3.5	4.9	
15	Nov.3, 04	9:26	HK	2.2	3.6	5.0	
16	Nov.3, 04	16:37	HK	2.2	3.6	5.1	
17	Nov.4, 04	7:36	HK	2.2	3.6	5.1	
18	Nov.4, 04	16:04	HK	2.2	3.6	5.1	
19	Nov.5, 04	9:37	HK	2.2	3.6	5.0	
20	Nov.5, 04	13:22	HK	2.2	3.6	5.1	
			Mean	2.2	3.6	5.0	Synchron CX Performance
			SD	0.0	0.1	0.1	SD = 0.3
			CV	0.0	2.8	2.0	CV = 4.5

Replication Experiment

- The CLIA criteria for acceptable performance states:
  - Short term = “within-run” or “within-day” experiment.
    - SD < 0.25 TE_a
  - Long- term = “between-day” experiment
    - SD < 0.33 TE_a

Method Validation

- Validation of Reportable Range or Linearity
- It is essential to assess the analytical range of a method, i.e., the lowest and highest test results that are reliable and can be reported.
- It is important to validate the manufacturer’s claims for reportable range of their system/method.

Reportable Range

- Factors to consider:
  - Sample selection
    - Standard solutions
    - Dilutions of a concentrated specimen
    - Proficiency Testing specimens for linearity
  - Use preferably 5 different levels of concentrations
    - May require more than 5 levels to determine where linearity “falls out”

Reportable Range Experiment

- Step 1: Prepare samples
  - Commercial samples or patient samples.
  - Choose at least 5 different concentrations
  - One near the zero level or estimated lower level of detection limit, and one slightly above the upper limit of the manufacturer's reportable range.

Reportable Range

- Step 2: Perform measurements
- NCCLS – 4 measurements on each specimen.
- Westgard – 3 measurements are sufficient.
- Calculate the mean of the measurements for each concentration level.

Reportable Range

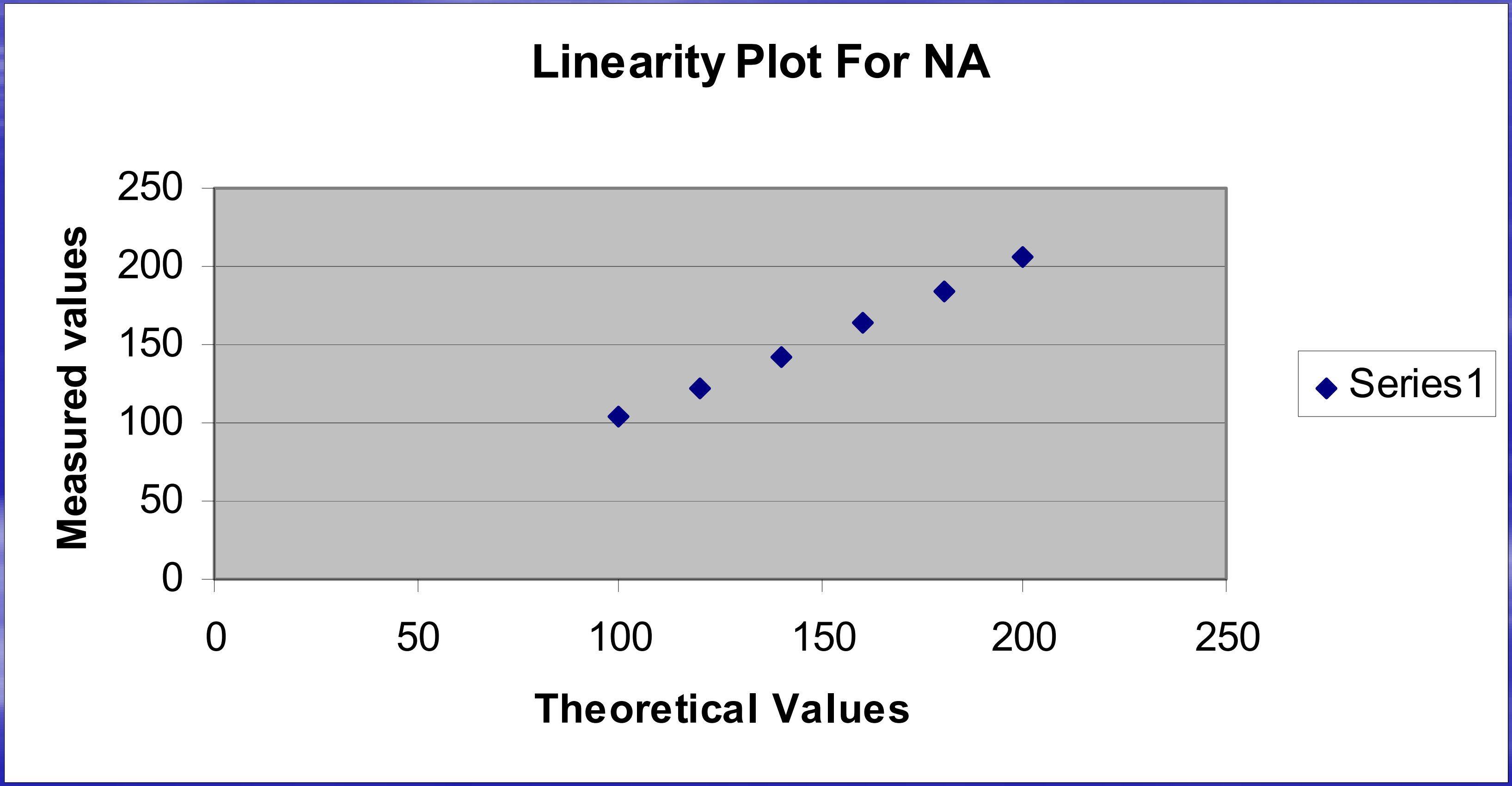
- Step 3: Plot data
- Measured mean values on y axis vs the known or assigned values on the x axis.
- Manually draw the best straight line through data points. (Do not use the computer)
  - Give more weight to the lowest points in the series.
- Inspect for linearity
- Make visual decision as to the acceptable reportable range.

Reportable Range – Example

- Analyte: Na   CX-5 Analytical Range: 100-200 mmol/L
- Lab Analytical Range: 100- **200 mmol/L**

Sample Level	Theoretical Value (x)	Measured Value Average (y)
M100 L-0/5	100	103.1
M100 L-1	120	122.1
M100 L-2	140	142
M100 L-3	160	164.1
M100 L-4	180	184.9
M100 L-5	200	205.6

Linearity Plot – Example 1



Method Validation

- Comparison of Methods:
- Performed to estimate inaccuracy or systematic error of the new method.
- Experiment is performed by analyzing patient samples by the new method (test method) and a comparative method, then estimate the systematic errors on the basis of the differences observed between the methods.

Comparison of Methods

- Comparative method:
  - Must be carefully selected, assumed to yield the correct results.
  - Any differences between a test method and a comparative method are assigned to the test method, because the correctness of the comparative method is well documented

Comparison of Methods-Measuring Inaccuracy

- Factors to consider:
  - Comparative method
    - Ideal = reference method
  - # of specimens to test
    - At least 40 patient samples
    - Cover the entire reportable range
    - One third in the low abnormal range, one third in the normal range and one third in the high abnormal range.
      - Use controls, standards or CAP survey material for spiking.

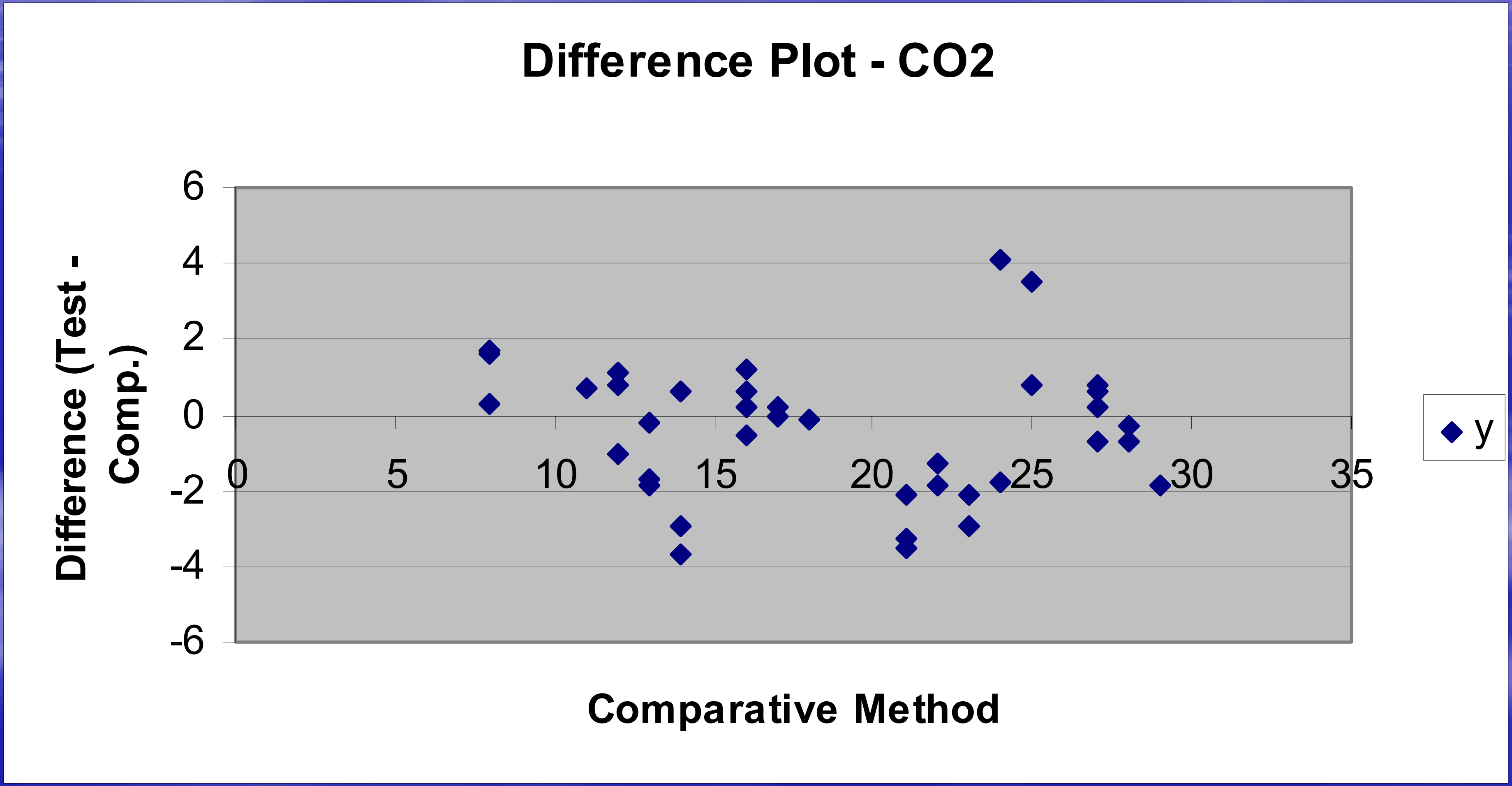
Comparison of Methods-Measuring Inaccuracy

- Single vs duplicate measurements
  - Sufficient Volume of specimen
- Time period
  - Test specimens on different days
  - Minimum 5 days, could extend to 20 days
  - Test the specimens on both methods simultaneously or within 2 hours of each other.

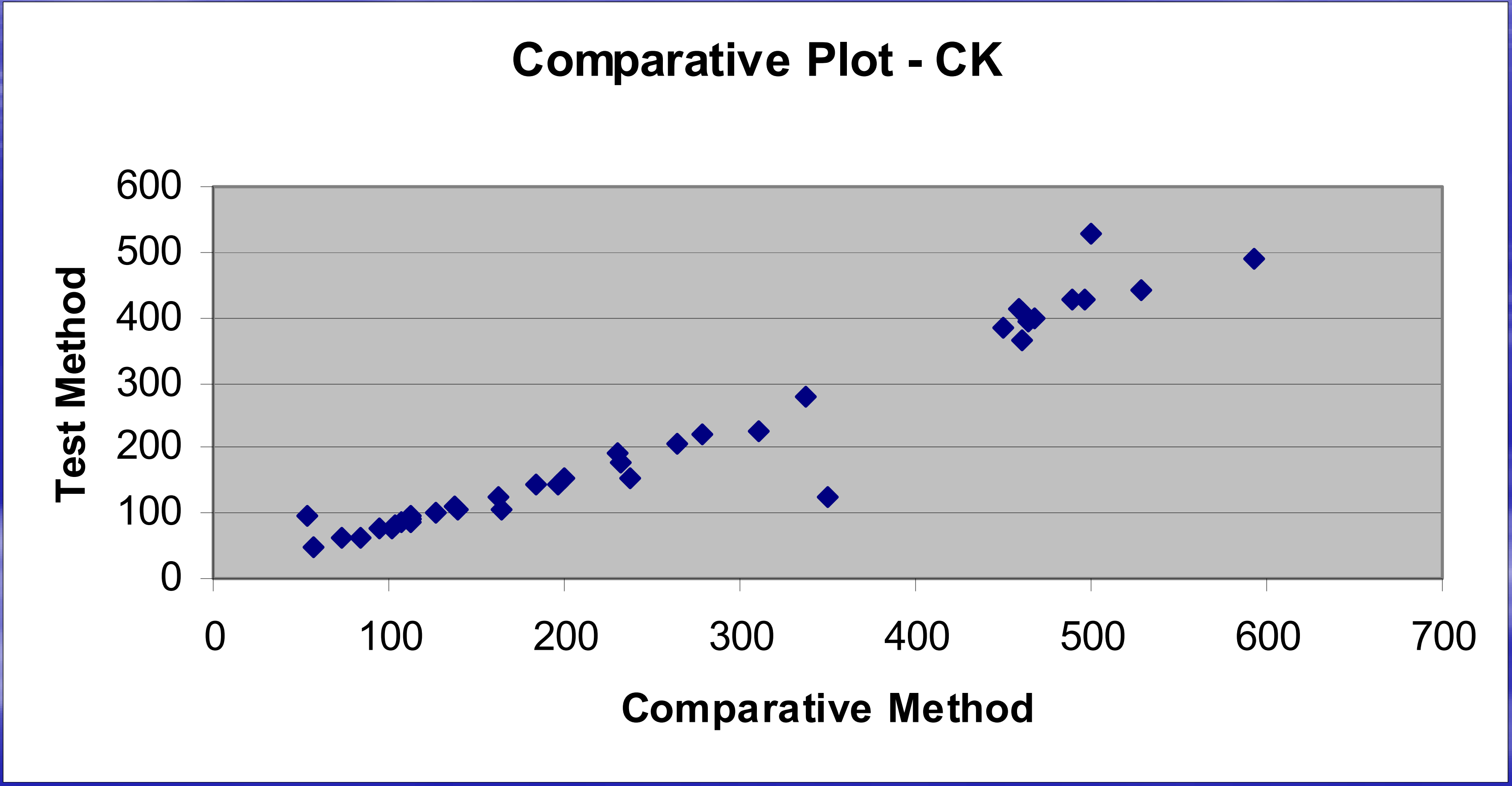
Comparison of Methods – Data Analysis

- 1. Graph the data:
  - Difference plot
    - Difference between the test results minus comparative results on y axis vs. comparative results on the x axis
    - Differences should scatter around the zero line.
    - Look for outliers and repeat the measurement.
  - Comparison plot
    - Plot the test values on the y axis vs the comparison values on the x axis.
    - Inspect for outliers and repeat.

Difference Plot - Example



Comparative Plot - Example



Comparison of Methods – Data Analysis

- 2. Calculate statistics:
- Different statistical tools are available for calculating the systematic error or bias.
- Linear regression analysis
- Paired t-test
- Bland Altman analysis
- Deming's regression
- Passing-Blalock regression
- Correlation Coefficient - r

Comparison of Methods - Statistics

- If you select to use linear regression statistics, you must evaluate the correlation coefficient (r):
- Correlation coefficient estimates the degree of association between two variables.
- If $r > 0.99$, use linear regression statistics
- If $r < 0.99$, use the paired t-test or another method.

Comparison of Methods

- Criteria for acceptable performance:
  - Must combine calculated random error (from the replication experiment) with the systematic error (from the comparison of methods experiment) to calculate TOTAL ERROR
  - TEcalc = SE + RE
  - TEcalc = bias + 3SD
  - TEcalc < TEa
  - Method performance is judged acceptable when the observed error (TEcalc is smaller than the defined allowable error (TEa)

Comparison of Methods

- USE A PROGRAM for statistical analysis!!!
- Westgard has a set of statistical tools on the internet to help calculate the statistics:
  - www.westgard.com/mvtools.html.
- Get a statistician from your data department to assist you.

Method Validation

- Reference Intervals or Normal Reference Range:
  - Verification of the manufacturer-supplied reference intervals for the population being served by the laboratory must be assessed.
  - It should be the last experiment to be studied in the method validation process.

Reference Intervals

- Several different ways to validate the transfer of the manufacturer's reference intervals to your individual lab.
- "divine judgement"
- Verification with 20 samples
- Estimation with 60 samples
- Full reference interval study

Reference Intervals

- Divine judgement
  - If there is consistency in demographics of the manufacturer’s study population and the population served by the local lab, then the manuf. reference intervals may be subjectively transferred to your lab.
  - Decision should be made by Lab Medical Director or equivalent.

Reference Intervals

- Verification with 20 samples
- To transfer the manufacturer's reference intervals to your lab:
- Test 20 samples from healthy individuals representing your local population.
- If < 3 values fall outside the manuf. reference interval, you may consider the reference interval verified.

Reference Intervals

- Estimation with 60 Samples:
- Collect and analyze samples from 60 healthy individuals from your local population.
- Estimate the reference intervals from the 60 samples and compare it with the reported manufacturer's intervals.

Reference Intervals

- Full Reference Interval Study
- Recommended when the demographics of the populations are different.
- Minimum requirement = 120 individuals from each group i.e. 120 men and 120 women.

Reference Intervals

- Client/Participant requirement:
- Selection accomplished by administering a health questionnaire.
- A consent form should signed by the participant, after counseling.
- If possible perform a physical examination.
- Screen individual for HIV.

Reference Intervals

- If HIV negative, draw blood for chemistry, hematology, CD4/CD8,and coagulation, HBsAg and HCV screening, etc.
- Perform the testing.
- Examine the data, exclude the outliers and HBsAg and HCV positive individuals.
- Perform statistical analysis on test results.
- Your new reference interval includes 95% (CI) of all your values i.e. Mean +/- 2SD.

Reference Intervals – UNC Project

- Total clients screened = 331
- HIV Positive Clients = 51
- Total enrolled as of Feb.9,2005 = 280
- Total tested for HBsAg = 234
- Total HBsAg Positive = 15 (6.4%)
- Total tested for HCV = 90
- Total HCV positive = 5 or (5.6%)

Method Validation- Additional Experiments

- Interference Experiment
- Detection Limit Experiment
- These are required for the modified non-waived tests.

Interference Experiment – Analytical Specificity

- Interference Experiment
  - Estimates systematic error caused by other materials that may be present in the specimen being analyzed.
    - e.g. lipemia, bilirubin, hemolysis etc
  - Compare the results between the neat specimen and the specimen with the added substance.

Detection Limit – Analytical Sensitivity

- Detection Limit Experiment
- Estimates the lowest concentration of an analyte that can be measured.
- Experiment performed by preparing a
- “blank” sample that has zero conc. of analyte
- “spiked” samples of low concentrations of analyte.
- Samples are measured repeatedly (replication),then the Means and SDs are calculated from the values obtained.

Method Validation -Summary

- Validation = estimating error

  Total error = systematic error + random error

- Essential components of MV:
  - Estimating imprecision (random error)
  - Estimating inaccuracy (systematic error)
  - Verifying reportable range (linearity)
  - Verifying reference intervals (normal reference range)

Summary

- For tests modified or developed in-house, one must also quantify:
- Analytical sensitivity (limit of detection)
- Analytical specificity (interfering substance)

Summary

Statistical Analysis

- Use available statistical tools.
- www.westgard.com/medxcel.htm
- www.westgard.com/mvtools.html

Summary

- 6 major points of MV:
  - Define quality requirement for your lab.
  - Select appropriate experiments
  - Collect experimental data
  - Perform statistics
  - Compare observed error with pre-determined allowable error (total allowable error)
  - Judge acceptability of observed method performance

Summary

- Make sure your experimental data is reviewed and signed by a lab supervisor or lab manager .
- Your data is filed neatly in a binder and is readily accessible to the monitors when they ask for it!

Old Method – Landmark AG II

NEW METHOD – BC CX-5 PRO